New measures of interest associated to chemical patterns: definition, implementation and experimental assessment

Leander Schietgat[1], Guillaume Poezevara[2], Ryan Bissell-Siders[2], Ronan Bureau[3], Bruno Crémilleux[2], Alban Lepailleur[3], Marie-Pierre Halm[3], Jan Ramon[1], Bertrand Cuissart[2]

[1]*Machine learning group, Department of Computer Science, Katholieke Universiteit Leuven*

[2]*CoDaG team, Groupe de Recherche en Informatique, Image, Automatique et Instrumentation de Caen (GREYC - UMR 6072), Université de Caen Basse-Normandie*

[3]*Centre d'Etudes et de Recherche sur le Médicament de Normandie (CERMN - UPRES EA 4258 - FR CNRS 3038 INC3M), Université de Caen Basse-Normandie.*

Abstract:

Chemistry today has to face a critical challenge, whose success necessitates high-performance computer tools all along the development process of a new molecule, from its earliest stages to its registration to the authorities. Our present work aims to develop such high-performance computer tools by automatically and inductively inferring general knowledge from chemical data. We have recently designed new methods for discovering association rules by directly mining the relational nature of the molecules as graphs. These methods link structural patterns with a given biochemical activity. All our methods rely on three measures of interest: the growth rate, the stimulation and the detection of scaffolds. First, given a set of compounds partitioned according to different biochemical behavior, the *growth rate of a structural pattern* corresponds to the odds linking the pattern with the different behavioral classes. We have designed a method that automatically extracts structural patterns with interesting growth rates from a set of molecular structures [1,2]. We then have successfully assessed this method on two critical problems: the prediction of the ecotoxicity and the prediction of the toxicity of a molecule [3]. Second, the *work on stimulation* corresponds to the design of a framework which takes into account not only the relations between structural patterns and their biochemical behavior, but also the context for the structural fragment [4]. Third, we have developed a method for *the detection of drug scaffolds*, which are basic building blocks of drug molecules. In this work, given a particular chemogenomics problem, the goal is to find interesting scaffolds that can help chemists to develop new drugs, possibly relating the scaffolds to the inhibition of a particular protein. All these results have been obtained within a framework obtained thanks to a long-term and close collaboration between the department of molecular modeling of the CERMN and the CoDaG team of the GREYC. This project has recently grown with recent and promising collaborations with members of the Machine Learning Group from the Katholieke Universiteit Leuven.

References:

(1) *Discovering emerging graph patterns from chemicals,* Guillaume Poezevara, Bertrand Cuissart and Bruno Crémilleux, in 18*th International Symposium on Methodologies for Intelligent Systems (ISMIS 2009),* Lecture Notes in Computer Science, pp 45-55, 2009.

(2) *Extracting and summarizing the frequent emerging graph patterns from a dataset of graphs***,** Guillaume Poezevara, Bertrand Cuissart, Bruno Crémilleux, to appear in Journal of Intelligent Information Systems**, 21 pages, 2011.**

(3) *Introduction of jumping fragments in combination with QSARs for the assessment of classification in ecotoxicology,* Sylvain Lozano, Guillaume Poezevara, Marie-Pierre Halm-Lemeille, Elodie Lescot-Fontaine, Alban Lepailleur, Ryan Bissell-Siders, Bruno Crémilleux, Sylvain Rault, Bertrand Cuissart and Ronan Bureau, *Journal of Chemical Information and Modeling,* vol. 50, pp. 1330-1339, 2010.

(4) O*n the stimulation of patterns: definitions, calculation method and first usages,* Ryan Bissell-Siders, Bertrand Cuissart and Bruno Crémilleux, in *Proc. of the 18th International Conference on Conceptual Structures (ICCS 2010)*, Lecture Notes in Artificial Intelligence, pp. 56-69, 2010.